ΑΝΑΛΥΣΗ ΔΙΑΣΠΟΡΑΣ, 24-5-02

Άσκηση 12. Τα παρακάτω δεδομένα προέρχονται από μετρήσεις του δείκτη του σακχάρου στο αίμα 30 ποντικών που εξετάσθηκαν: 1) υπό κανονικές συνθήκες, 2) μετά από ένεση pitressin, 3) μετά από ένεση pitocin (στα ποντίκια δεν επιδρούν άλλοι παράγοντες π.χ. ηλικία, βάρος, φύλο κ.τ.λ.)

1	93	111	118	98	114	107	108	120	106	96	114	125
2	134	129	121	127	138	126	134	130	147	135		
3	118	106	115	95	109	108	104	108				

1) Κατασκευάστε ένα scatterplot και ένα Boxplot των δεδομένων.
2) Εκτιμήστε τα μ_1, μ_2, μ_3 σημειακά και με δ.ε. 95% (χρησιμοποιώντας individual variance estimate και pooled variance estimate).
3) Να κατασκευάσετε τα γραφήματα των παραπάνω δ.ε.
4) Εξετάστε την υπόθεση H_0: $\mu_1 = \mu_2 = \mu_3$ με H_1: όχι $\mu_1 = \mu_2 = \mu_3$ σε ε.σ. $\alpha = 1\%$.
5) Βρείτε 95% δ.ε. για τη διαφορά $\mu_2 - \mu_1$ και 95% δ.ε. για τη διαφορά $\mu_3 - \mu_1$.
6) Να δοθούν ταυτόχρονα δ.ε. για τις διαφορές $\mu_2 - \mu_1, \mu_3 - \mu_1, \mu_3 - \mu_2$ με τις μεθόδους a) Bonferroni b) Tukey, c) Sheffe. Να γίνει ομαδοποίηση των μέσων μ_1, μ_2, μ_3 με πιθανότητα σφάλματος 5%.
7) Να ελέγξετε αν οι διασπορές των παρατηρήσεων είναι ίδιες στις περιπτώσεις 1, 2, 3 με $\alpha = 0.01$.
8) Να εκτιμήσετε τα μ_1, μ_2, μ_3 και τον πίνακα ANOVA χρησιμοποιώντας την ανάλυση παλινδρόμησης.

Άσκηση 13. Εκλέχθηκαν τυχαία 50 δείγματα από πέντε τύπους αυτοκινήτων (1,2,3,4,5) και μετρήθηκε η κατανάλωση βενζίνας σε κάθε περίπτωση (κάτω από τις ίδιες συνθήκες κυκλοφορίας). Τα αποτελέσματα (σε χιλιόμετρα ανά γαλόνι) ήταν:

τύπος													
1	13,0	15,2	21,2	21,4	21,7	17,7	19,3	16,5	16,1	14,6	17,7		
2	25,4	27,4	29,1	30,1	25,6	23,5	29,2	29,8					
3	16,6	23,0	20,6	19,9	22,0	25,2	16,7	25,2	24,5	27,1			
4	27,6	24,4	28,9	32,8	20,3	27,8	22,4	23,6					
5	31,8	32,3	31,1	27,9	30,2	35,7	32,1	38,5	30,6	28,5	29,6	34,9	36,6

1) Κατασκευάστε scatterplot και Boxplot των δεδομένων.
2) Να ελέγξετε αν οι διασπορές των παρατηρήσεων είναι ίδιες στις περιπτώσεις 1, 2, 3, 4, 5 με α = 0.05.
3) Εκτιμήστε τα $\mu_1,..., \mu_5$ σημειακά και με δ.ε. 95% (χρησιμοποιώντας individual και pooled variance estimates). Να κατασκευάσετε τα γραφήματα των συγκεκριμένων δ.ε.
4) Εξετάστε την υπόθεση H_0: $\mu_1 = ... = \mu_5$ με H_1: όχι $\mu_1 = ... = \mu_5$ σε ε.σ. α = 1%.
5) Βρείτε 95% δ.ε. για τη διαφορά $\mu_5 - \mu_1$ και 95% δ.ε. για τη διαφορά $\mu_4 - \mu_2$ και εξετάστε αν $\mu_4 = \mu_2$ (με εναλλακτική $\mu_4 \neq \mu_2$) σε ε.σ. 1%.
6) Να δοθούν ταυτόχρονα δ.ε. για όλες τις ανά δύο διαφορές $\mu_i - \mu_j$ με τις μεθόδους a) Bonferroni b) Tukey, c) Sheffe. Κατατάξτε τα $\mu_1, \mu_2, ..., \mu_5$ με πιθανότητα σφάλματος 5%.
7) Να εκτιμήσετε τα $\mu_1, \mu_2, ..., \mu_5$ και τον πίνακα ANOVA χρησιμοποιώντας την ανάλυση παλινδρόμησης.

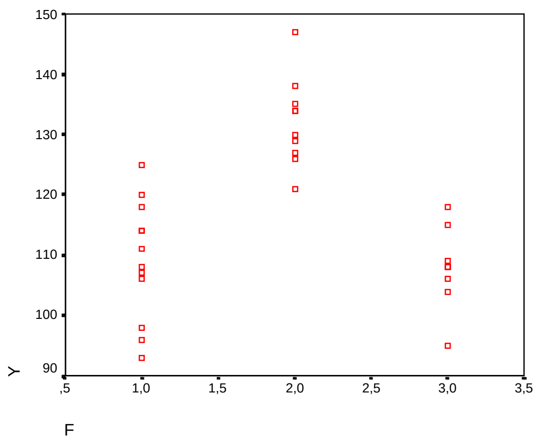
Απαντήσεις Άσκησης 12.

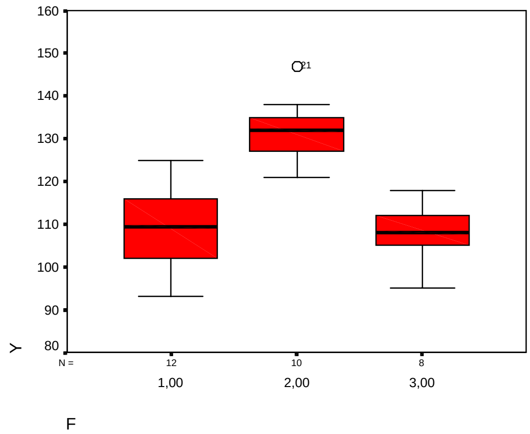
Αρχικά εισάγουμε τις 30 περιπτώσεις στο SPSS σε τρεις στήλες *i*, *y*, *f* (αρκούν και οι στήλες *y*, *f*) ως εξής:

i	***y***	***f***
1	93	1
2	111	1
3	118	1
4	98	1
5	114	1
6	107	1
7	108	1
8	120	1
9	106	1
10	96	1
11	114	1
12	125	1
13	134	2
14	129	2
15	121	2

i	***y***	***f***
16	127	2
17	138	2
18	126	2
19	134	2
20	130	2
21	147	2
22	135	2
23	118	3
24	106	3
25	115	3
26	95	3
27	109	3
28	108	3
29	104	3
30	108	3

1) Για scatterplot: Graphs/Scatterplot/simple, Y axis: y, X axis: f. Για Boxplot: Graphs/Boxplot/Simple: Variable:y, Category Axis: f





Παρατηρούμε ότι οι τιμές των δεδομένων της κατηγορίας 2 (μετρήσεις του δείκτη του σακχάρου μετά από ένεση pitressin) φαίνονται να είναι μεγαλύτερες από τις τιμές των κατηγοριών 1, 3.

2) Εφαρμόζουμε το μοντέλο

$$Y_{ij} = \mu_i + \varepsilon_{ij},\ i = 1,2,3\ \ (r = 3), \quad j = 1,2,\ldots,n_i,$$

όπου ε_{ij} ανεξάρτητες τ.μ. που ακολουθούν $N(0,\sigma^2)$.

Στο SPSS: Analyze/compare means/one-way ANOVA, Dependent list: y Factor: f, Options: Check Descriptive λαμβάνεται ο πίνακας:

Descriptives

Y

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean: Lower Bound	95% Confidence Interval for Mean: Upper Bound	Minimum	Maximum
1,00	12	109,1667	9,8704	2,8493	102,8953	115,4380	93,00	125,00
2,00	10	132,1000	7,2488	2,2923	126,9146	137,2854	121,00	147,00
3,00	8	107,8750	6,9578	2,4599	102,0581	113,6919	95,00	118,00
Total	30	116,4667	13,8433	2,5274	111,2975	121,6358	93,00	147,00

Ο οποίος περιέχει:

	N	Mean	Std. Dev.	Std. Error	Lower/Upper Bound
1	n_1	$\bar{Y}_{1\bullet}$	$S_1=\sqrt{\frac{1}{n_1-1}\sum_{j=1}^{n_1}(Y_{1j}-\bar{Y}_{1\bullet})^2}$	$s(\bar{Y}_{1\bullet})=\frac{S_1}{\sqrt{n_1}}$	$\bar{Y}_{1\bullet}\pm\frac{S_1}{\sqrt{n_1}}t_{n_1-1}(\frac{a}{2})$
2	n_2	$\bar{Y}_{2\bullet}$	$S_2=\sqrt{\frac{1}{n_2-1}\sum_{j=1}^{n_2}(Y_{2j}-\bar{Y}_{2\bullet})^2}$	$s(\bar{Y}_{2\bullet})=\frac{S_2}{\sqrt{n_2}}$	$\bar{Y}_{2\bullet}\pm\frac{S_2}{\sqrt{n_2}}t_{n_2-1}(\frac{a}{2})$
3	n_3	$\bar{Y}_{3\bullet}$	$S_3=\sqrt{\frac{1}{n_3-1}\sum_{j=1}^{n_3}(Y_{3j}-\bar{Y}_{3\bullet})^2}$	$s(\bar{Y}_{3\bullet})=\frac{S_3}{\sqrt{n_3}}$	$\bar{Y}_{3\bullet}\pm\frac{S_3}{\sqrt{n_3}}t_{n_3-1}(\frac{a}{2})$
Total	n_T	$\bar{Y}_{\bullet\bullet}$	$S_T=\sqrt{\frac{1}{n_T-1}\sum_{i,j}(Y_{ij}-\bar{Y}_{\bullet\bullet})^2}$	$s(\bar{Y}_{\bullet\bullet})=\frac{S_T}{\sqrt{n_T}}$	$\bar{Y}_{\bullet\bullet}\pm\frac{S_T}{\sqrt{n_T}}t_{n_T-1}(\frac{a}{2})$

(όπου α = 0.05)

Επίσης λαμβάνεται και ο πίνακας

ANOVA

Y

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3674,025	2	1837,012	26,334	,000
Within Groups	1883,442	27	69,757		
Total	5557,467	29			

ο οποίος περιέχει:

Model	***Sum of squares***	***df***	***Mean Square***	***F***	***Sig.***
Between Groups	$SSTr=\sum_i n_i(\bar{Y}_{i\bullet}-\bar{Y}_{\bullet\bullet})^2\left(=\sum(\hat{Y}_i-\bar{Y})^2\right)$	$r-1$	$MSTr=\frac{SSR}{r-1}$	$F^*=\frac{MSR}{MSE}$	$P(F>F^* \mid F\sim F_{r-1,n-r})$
Within Groups	$SSE=\sum_{i,j}(Y_{ij}-\bar{Y}_{i\bullet})^2\left(=\sum(Y_i-\hat{Y}_i)^2\right)$	n_T-r	$MSE=\frac{SSE}{n_T-r}$		
Total	$SST=\sum_{i,j}(Y_{ij}-\bar{Y}_{\bullet\bullet})^2\left(=\sum(Y_i-\bar{Y})^2\right)$	n_T-1			

Προφανώς, τα παραπάνω δ.ε. προκύπτουν χρησιμοποιώντας internal variance estimates (οι εκτιμήσεις των διασπορών γίνονται σε κάθε ομάδα ξεχωριστά). Η pooled εκτίμηση της διασποράς (χρησιμοποιείται όταν οι διασπορές μέσα στις ομάδες θεωρούνται ίσες) είναι η

$$s^2=MSE=\frac{SSE}{n_T-r}=\frac{1}{n_T-r}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i\bullet})^2=\frac{1}{n_T-r}\sum_{i=1}^{r}(n_i-1)S_i^2=69.757$$

και άρα τα δ.ε. 1–α για τα μ_i με pooled εκτίμηση της διασποράς είναι

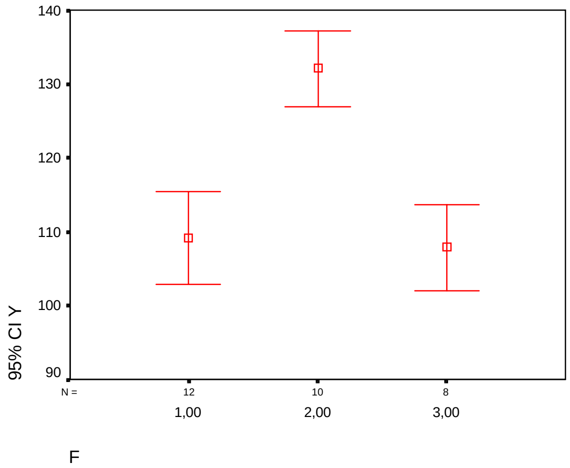
$$\bar{Y}_{1\bullet}\pm\frac{s}{\sqrt{n_1}}t_{n_T-r}(\frac{a}{2})=109.1667\pm\frac{\sqrt{69.757}}{\sqrt{12}}t_{27}(0.025)=(104.22,114.11)$$

$$\bar{Y}_{2\bullet}\pm\frac{s}{\sqrt{n_2}}t_{n_T-r}(\frac{a}{2})=132.1\pm\frac{\sqrt{69.757}}{\sqrt{10}}t_{27}(0.025)=(126.68,137.51)$$

$$\bar{Y}_{3\bullet} \pm \frac{s}{\sqrt{n_3}} t_{n_T - r}(\frac{a}{2}) = 107.875 \pm \frac{\sqrt{69.757}}{\sqrt{8}} t_{27}(0.025) = (101.82, 113.93)$$

όπου $t_{27}(0.025) = IDF.T(0.975,27) \approx 2.05$ (υπολογίζονται χρησιμοποιώντας π.χ. το transform / compute).

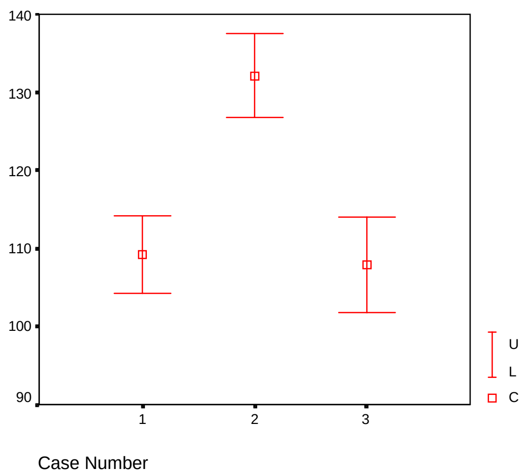
3) Επιλέγουμε Graphs/error bars/simple, variable: y, Category Axis: f (c.i. for mean, level: 95%) από όπου λαμβάνουμε το γράφημα με τα δ.ε. 95% για τα μ_i (individual variance estimates).



Για να φτιάξουμε το αντίστοιχο γράφημα με τα pooled estimates θα εργασθούμε ως εξής: φτιάχνουμε τρεις μεταβλητές (μπορούμε προσωρινά να κλείσουμε τα δεδομένα της άσκησης) που περιέχουν τα συγκεκριμένα δ.ε. (με τα κεντρικά σημεία) ως εξής:

L	*C*	*U*
104.22	109.16	114.11
126.68	132.1	137.51
101.82	107.87	113.93

και επιλέγουμε Graphs/High-Low/Simple High-Low, check values of individual cases, High:U, Low: L, close: C από όπου λαμβάνουμε το γράφημα με τα δ.ε. 95% για τα μ_i (pooled variance estimates):



Παρατηρούμε ότι το δ.ε. για το μ_2 δεν τέμνεται με τα άλλα δύο δ.ε. άρα υπάρχει μία ένδειξη ότι το μ_2 μπορεί να διαφέρει από τα μ_1, μ_3 (ο ακριβής έλεγχος γίνεται παρακάτω).

4) Ο έλεγχος της υπόθεσης H_0: $\mu_1 = \mu_2 = \mu_3$ έναντι της H_1: όχι $\mu_1 = \mu_2 = \mu_3$ γίνεται χρησιμοποιώντας το *F - test* που περιέχεται στον πίνακα ANOVA. Παρατηρούμε ότι το αντίστοιχο p-value είναι 0.00

(F^* = 26.334) οπότε απορρίπτουμε ότι $\mu_1 = \mu_2 = \mu_3$ σε ε.σ. 1% (η μέση τιμή δεν είναι ίση και στις τρεις ομάδες).

5) Επιλέγουμε Analyze/compare means/one-way ANOVA, Dependent list: y Factor: f, : Post Hoc: LSD απ' όπου λαμβάνουμε:

Multiple Comparisons

Dependent Variable: Y
LSD

(I) F	(J) F	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1,00	2,00	-22,9333*	3,5761	,000	-30,2710	-15,5957
	3,00	1,2917	3,8122	,737	-6,5303	9,1136
2,00	1,00	22,9333*	3,5761	,000	15,5957	30,2710
	3,00	24,2250*	3,9617	,000	16,0962	32,3538
3,00	1,00	-1,2917	3,8122	,737	-9,1136	6,5303
	2,00	-24,2250*	3,9617	,000	-32,3538	-16,0962

*. The mean difference is significant at the .05 level.

Ο παραπάνω πίνακας περιέχει (s^2 = MSE = pooled varince estimate):

i	j	$\hat{\mu}_i - \hat{\mu}_j$	$s.e.$	sig	LB, UB
1	2	$\bar{Y}_{1\bullet} - \bar{Y}_{2\bullet}$	$s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	$P(\|T\| > \left\|\frac{\bar{Y}_{1\bullet} - \bar{Y}_{2\bullet}}{s.e.}\right\| \mid T \sim t_{n_T-r})$	$\bar{Y}_{1\bullet} - \bar{Y}_{2\bullet} \pm t_{n_T-r}(\frac{a}{2}) \cdot s.e.$
	3	$\bar{Y}_{1\bullet} - \bar{Y}_{3\bullet}$	$s\sqrt{\frac{1}{n_1}+\frac{1}{n_3}}$	$P(\|T\| > \left\|\frac{\bar{Y}_{1\bullet} - \bar{Y}_{3\bullet}}{s.e.}\right\| \mid T \sim t_{n_T-r})$	$\bar{Y}_{1\bullet} - \bar{Y}_{3\bullet} \pm t_{n_T-r}(\frac{a}{2}) \cdot s.e.$
...	...	...	...	...	...

οπότε το 95% δ.ε. για τη διαφορά $\mu_2 - \mu_1$ είναι (15.59, 30.27) και το 95% δ.ε. για τη διαφορά $\mu_3 - \mu_1$ είναι (–9.11, 6.53). Επειδή το δ.ε. για το $\mu_2 - \mu_1$ δεν περιέχει το 0 μπορούμε να πούμε ότι το μ_1 διαφέρει σημαντικά από το μ_2 (ε.σ. 5%).

Οι έλεγχοι μέσω των παραπάνω δ.ε. είναι σε ε.σ. 5% **ο καθένας**, δηλαδή π.χ. P(*σφάλμα τύπου I*) = P(*απορρίπτουμε την* H_0: $\mu_1 = \mu_2$ | *ισχύει η* H_0) = 5%. Επομένως, η πιθανότητα να γίνει κάποια λάθος απόρριψη στους τρεις ελέγχους $\mu_1 = \mu_2$, $\mu_2 = \mu_3$, $\mu_1 = \mu_3$ είναι μεγαλύτερη του 5% διότι αν A_{ij} = {*απορρίπτουμε λανθασμένα την* H_0: $\mu_i = \mu_j$ } τότε

$$P(A_{12} \cup A_{13} \cup A_{23}) = 1 - P(A_{12}^c \cap A_{13}^c \cap A_{23}^c) \approx 1 - P(A_{12}^c)P(A_{13}^c)P(A_{23}^c) = 1-(1-0.05)^3 \approx 14.2\%$$

(η ισότητα παραπάνω ισχύει αν τα ενδεχόμενα A_{ij} είναι ανεξάρτητα). Επομένως, για πολλαπλές συγκρίσεις (όπου η πιθανότητα να κάνουμε κάποια λάθος απόρριψη σε όλους τους ελέγχους που πραγματοποιούμε είναι το πολύ a) θα πρέπει να χρησιμοποιήσουμε κάποια από τις μεθόδους Tukey, Sheffe, Bonferroni.

6) Επιλέγουμε Analyze/compare means/one-way ANOVA, Dependent list: y Factor: f, : Post Hoc: Tukey, Sheffe, Bonferroni απ' όπου λαμβάνουμε:

Multiple Comparisons

Dependent Variable: Y

	(I) F	(J) F	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Tukey HSD	1,00	2,00	-22,9333*	3,5761	,000	-31,8001	-14,0666
		3,00	1,2917	3,8122	,939	-8,1603	10,7437
	2,00	1,00	22,9333*	3,5761	,000	14,0666	31,8001
		3,00	24,2250*	3,9617	,000	14,4022	34,0478
	3,00	1,00	-1,2917	3,8122	,939	-10,7437	8,1603
		2,00	-24,2250*	3,9617	,000	-34,0478	-14,4022
Scheffe	1,00	2,00	-22,9333*	3,5761	,000	-32,1957	-13,6710
		3,00	1,2917	3,8122	,944	-8,5820	11,1653
	2,00	1,00	22,9333*	3,5761	,000	13,6710	32,1957
		3,00	24,2250*	3,9617	,000	13,9640	34,4860
	3,00	1,00	-1,2917	3,8122	,944	-11,1653	8,5820
		2,00	-24,2250*	3,9617	,000	-34,4860	-13,9640
Bonferroni	1,00	2,00	-22,9333*	3,5761	,000	-32,0613	-13,8054
		3,00	1,2917	3,8122	1,000	-8,4388	11,0221
	2,00	1,00	22,9333*	3,5761	,000	13,8054	32,0613
		3,00	24,2250*	3,9617	,000	14,1128	34,3372
	3,00	1,00	-1,2917	3,8122	1,000	-11,0221	8,4388
		2,00	-24,2250*	3,9617	,000	-34,3372	-14,1128

*. The mean difference is significant at the .05 level.

α) Όσον αφορά τη μέθοδο Bonferroni ο παραπάνω πίνακας αποτελείται:

i	j	$\hat{\mu}_i - \hat{\mu}_j$	*s.e.*	*sig*	*LB, UB*
i	j	$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}$	$s\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$	$\min\{1,\ 3P(\lvert T\rvert > \left\lvert\frac{\bar{Y}_{i\bullet}-\bar{Y}_{j\bullet}}{s.e.}\right\rvert \mid T \sim t_{n_T-r})\}$	$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} \pm t_{n_T-r}(\frac{a/3}{2}) \cdot s.e.$

διότι αν $P(A_{ij}) = P(\text{απορρίπτουμε λανθασμένα την } H_0\text{: } \mu_i = \mu_j) = a/r$ τότε

$$P(A_{12} \cup A_{13} \cup A_{23}) \le P(A_{12}) + P(A_{13}) + P(A_{23}) = a/3 + a/3 + a/3 = a$$

Δηλαδή, η πιθανότητα να απορρίψουμε λανθασμένα κάποια από τις υποθέσεις είναι το πολύ α.

Όσον αφορά τη μέθοδο Tukey, ο παραπάνω πίνακας αποτελείται:

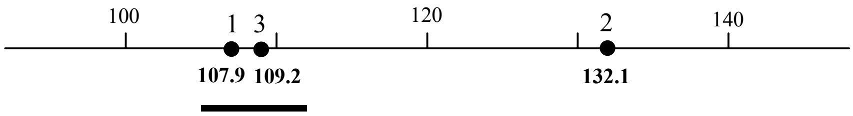
i	j	$\hat{\mu}_i - \hat{\mu}_j$	*s.e.*	*sig*	*LB, UB*
i	j	$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}$	$s\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$	$P(\lvert T\rvert > \sqrt{2}\left\lvert\frac{\bar{Y}_{i\bullet}-\bar{Y}_{j\bullet}}{s.e.}\right\rvert \mid T \sim q(r, n_T - r))$	$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} \pm \frac{q(a; r, n_T - r)}{\sqrt{2}} \cdot s.e.$

Υπενθυμίζεται ότι $q(a; r, v)$ είναι το άνω a-σημείο της κατανομής $q(r,v)$ του τυποποιημένου εύρους $\frac{X_{(r)} - X_{(1)}}{s}$ (όπου $X_1, \ldots, X_r$ ανεξ. τ.μ. από $N(0,\sigma^2)$ και $vs^2/\sigma^2 \sim \chi^2_v$ ανεξ. των X_i).

Τέλος, όσον αφορά τη μέθοδο Sheffe, ο παραπάνω πίνακας αποτελείται:

i	j	$\hat{\mu}_i - \hat{\mu}_j$	*s.e.*	*sig*	*LB, UB*
i	j	$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}$	$s\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$	$P(T > \frac{1}{r-1}\left(\frac{\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}}{s.e.}\right)^2 \mid T \sim F_{r-1,n_T-r})$	$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} \pm \sqrt{(r-1)F_{r-1,n_T-r}(a)} \cdot s.e.$

Παρατηρούμε ότι σε σύμφωνα με όλες τις παραπάνω μεθόδους, απορρίπτουμε ότι $\mu_1 = \mu_2$, $\mu_2 = \mu_3$ ενώ δεχόμαστε ότι $\mu_1 = \mu_3$ σε ε.σ. 5%. Δηλαδή, το μ_2 διαφέρει από τα μ_1, μ_3



(Δηλαδή η ένεση pitressin (2) έχει σημαντική επίδραση στο δείκτη του σακχάρου ενώ η ένεση pitocin (3) δεν φαίνεται να έχει σημαντική επίδραση). Η παραπάνω ομαδοποίηση δίνεται και από το πακέτο:

Y

	F	N	Subset for alpha = .05: 1	2
Tukey HSD[a,b]	3,00	8	107,8750	
	1,00	12	109,1667	
	2,00	10		132,1000
	Sig.		,938	1,000
Scheffe[a,b]	3,00	8	107,8750	
	1,00	12	109,1667	
	2,00	10		132,1000
	Sig.		,944	1,000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 9,730.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Αξίζει να παρατηρηθεί ότι για την εκτίμηση γραμμικών αντιθέσεων της μορφής $c_1\mu_1+c_2\mu_2+...+c_r\mu_r$ όπου c_i γνωστοί συντελεστές με $c_1+c_2+...+c_r = 0$, μπορούμε να χρησιμοποιήσουμε την επιλογή contrasts στη διαδικασία one-way ANOVA του SPSS. Συγκεκριμένα εισάγουμε ένα-ένα τα c_i στο πεδίο coefficients (πατώντας κάθε φορά add).

7) Για να ελέγξουμε αν οι διασπορές των παρατηρήσεων είναι ίδιες στις περιπτώσεις 1, 2, 3 με α = 0.1 θα χρησιμοποιήσουμε το τεστ Levene. Ο έλεγχος αυτός θα πρέπει να γίνεται στην αρχή της ανάλυσης για να βεβαιωθούμε ότι το μοντέλο στο οποίο βασιζόμαστε είναι σωστό (οι διασπορές των ε_{ij} είναι ίσες για όλες τις ομάδες).

Επιλέγουμε Analyze/compare means/one-way ANOVA, Dependent list: y Factor: f, : Options: check Homogeneity of variance από όπου παίρνουμε

Test of Homogeneity of Variances

Y

Levene Statistic	df1	df2	Sig.
1,126	2	27	,339

Υπενθυμίζεται ότι το Levene τεστ γίνεται ως εξής: Δημιουργούνται οι απόλυτες διαφορές $X_{ij} = | Y_{ij} - \overline{Y}_{i\bullet} |$, i=1,2,...,r, j = 1, 2, ..., n_i και εκτελείται ανάλυση διασποράς στις διαφορές αυτές (dependent:X, Factor: f). Το Levene statistic είναι το *F-ratio* του πίνακα ANOVA που προκύπτει. Αντίστοιχα, το παραπάνω p-value (=0.339) είναι το p-value του F-test στο συγκεκριμένο πίνακα ANOVA. Ο έλεγχος αυτός είναι περισσότερο ευσταθής από άλλα παρόμοια τεστ (Bartlett, Cochran, Hartley) στην περίπτωση μη-κανονικότητας των παρατηρήσεων.

Εφ’όσον το p-value = 0.339 δεχόμαστε ότι οι διασπορές των παρατηρήσεων είναι ίδιες στις περιπτώσεις 1, 2, 3.

8) Είναι γνωστό ότι το μοντέλο μας μπορεί εναλλακτικά να γραφεί στη μορφή $Y = \mathbf{X}\beta + \varepsilon$ όπου

$$Y = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \vdots \\ \vdots \\ Y_{r1} \\ \vdots \\ Y_{rn_r} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & & \\ \vdots & & \\ 1 & & \\ & \ddots & \\ & & \ddots \\ & & 1 \\ & & \vdots \\ & & 1 \end{bmatrix}, \beta = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_r \end{bmatrix} \varepsilon = \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \vdots \\ \vdots \\ \varepsilon_{r1} \\ \vdots \\ \varepsilon_{rn_r} \end{bmatrix}$$

Συνεπώς, κατασκευάζουμε τις μεταβλητές:

y	x1	x2	x3
93	1	0	0
111	1	0	0
118	1	0	0
98	1	0	0
114	1	0	0
107	1	0	0
108	1	0	0
120	1	0	0
106	1	0	0
96	1	0	0
114	1	0	0
125	1	0	0
134	0	1	0
129	0	1	0
121	0	1	0
127	0	1	0
138	0	1	0
126	0	1	0
134	0	1	0
130	0	1	0
147	0	1	0
135	0	1	0
118	0	0	1
106	0	0	1
115	0	0	1
95	0	0	1
109	0	0	1
108	0	0	1
104	0	0	1
108	0	0	1

και επιλέγουμε Analyze/Regression/Linear/Dependent:*y*, Independent:X_1,X_2,X_3, options**: do not include constant in equation**, Statistics: check confidence intervals από όπου λαμβάνουμε τον πίνακα:

Coefficients[a,b]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	X1	109,167	2,411	,589	45,278	,000	104,220	114,114
	X2	132,100	2,641	,650	50,016	,000	126,681	137,519
	X3	107,875	2,953	,475	36,532	,000	101,816	113,934

a. Dependent Variable: Y

b. Linear Regression through the Origin

ο οποίος αποτελείται από:

Model	B	Std. Error (pooled)	t	sig.	Lower/Upper Bound
X_1	$\bar{Y}_{1\bullet}$	$\frac{s}{\sqrt{n_1}}$	$T^* = \frac{\bar{Y}_{1\bullet}}{s/\sqrt{n_1}} \left(= \frac{\hat{\beta}_1}{s(\hat{\beta}_1)} \right)$	$P(\lvert T \rvert > \lvert T^* \rvert \mid T \sim t_{n_T - r})$	$\bar{Y}_{1\bullet} \pm \frac{s}{\sqrt{n_1}} t_{n_T - r}(\frac{a}{2})$
X_2	$\bar{Y}_{2\bullet}$	$\frac{s}{\sqrt{n_2}}$	$T^* = \frac{\bar{Y}_{2\bullet}}{s/\sqrt{n_2}} \left(= \frac{\hat{\beta}_2}{s(\hat{\beta}_2)} \right)$	$P(\lvert T \rvert > \lvert T^* \rvert \mid T \sim t_{n_T - r})$	$\bar{Y}_{2\bullet} \pm \frac{s}{\sqrt{n_2}} t_{n_T - r}(\frac{a}{2})$
X_3	$\bar{Y}_{3\bullet}$	$\frac{s}{\sqrt{n_3}}$	$T^* = \frac{\bar{Y}_{3\bullet}}{s/\sqrt{n_3}} \left(= \frac{\hat{\beta}_3}{s(\hat{\beta}_3)} \right)$	$P(\lvert T \rvert > \lvert T^* \rvert \mid T \sim t_{n-2})$	$\bar{Y}_{3\bullet} \pm \frac{s}{\sqrt{n_3}} t_{n_T - r}(\frac{a}{2})$

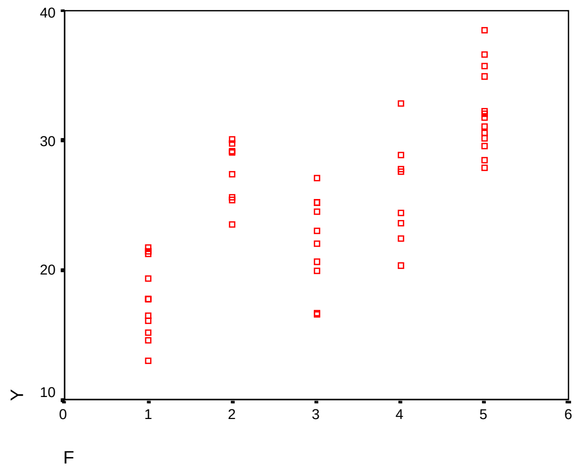
όπου s^2 = MSE είναι η pooled εκτίμηση του σ^2 και α=0.05. Παρατηρούμε ότι τα παραπάνω δ.ε. είναι ακριβώς αυτά που υπολογίσαμε στο (2) χρησιμοποιώντας pooled variance estimates.

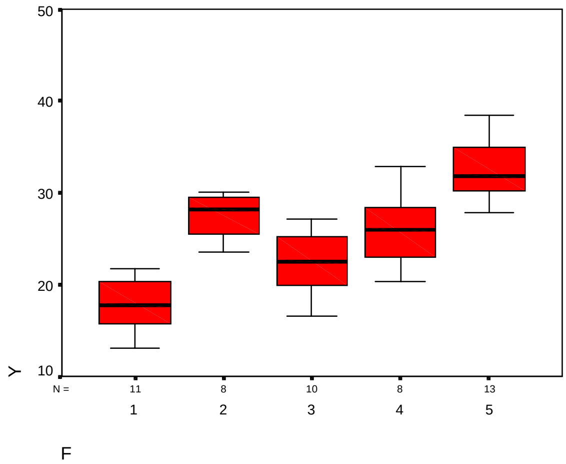
Τέλος, ο πίνακας ANOVA που δίνεται από την παραπάνω ανάλυση παλινδρόμησης παρατηρούμε ότι δεν είναι ίδιος με τον πίνακα ANOVA που βρήκαμε στο (2) διότι σε αυτή την περίπτωση τα αθροίσματα τετραγώνων SST και SSR δεν «διορθώνονται» με αφαίρεση του $\bar{Y}$ (δηλ. $SST = \sum Y_i^2$, $SSR = \sum \hat{Y}_i^2$) (αυτό γίνεται από το πακέτο γιατί έχουμε υποθέσει ότι $\beta_0 = 0$ εξαιρώντας τη σταθερά από την ανάλυση). Για να λάβουμε τον σωστό πίνακα θα πρέπει να συμπεριλάβουμε τη σταθερά στο μοντέλο. Σε αυτή την περίπτωση δεν θα πρέπει να βάλουμε όλες τις στήλες X_1,X_2,X_3 διότι ο πίνακας σχεδιασμού $\mathbf{X} = (\mathbf{1}, X_1, X_2, X_3)$ θα περιέχει γραμμικά εξαρτημένες στήλες ($\mathbf{1} = X_1+X_2+X_3$). Αρκεί να βάλουμε π.χ. τις στήλες X_2, X_3. (Σε αυτή την περίπτωση ο πίνακας των Coefficients θα περιέχει εκτιμήσεις για τα $\beta_0 = \mu_1$, $\beta_1 = \mu_2 - \mu_1$, $\beta_2 = \mu_3 - \mu_1$ και τα δ.ε. θα είναι τα αντίστοιχα LSD ε.σ. 95%).

Απαντήσεις Άσκησης 13. Αρχικά εισάγουμε τα δεδομένα στο SPSS (2 μεταβλητές με 50 γραμμές):

y	*f*	*y*	*f*	*y*	*f*	*y*	*f*	*y*	*f*
13	1	17,7	1	23	3	24,4	4	27,9	5
15,2	1	25,4	2	20,6	3	28,9	4	30,2	5
21,2	1	27,4	2	19,9	3	32,8	4	35,7	5
21,4	1	29,1	2	22	3	20,3	4	32,1	5
21,7	1	30,1	2	25,2	3	27,8	4	38,5	5
17,7	1	25,6	2	16,7	3	22,4	4	30,6	5
19,3	1	23,5	2	25,2	3	23,6	4	28,5	5
16,5	1	29,2	2	24,5	3	31,8	5	29,6	5
16,1	1	29,8	2	27,1	3	32,3	5	34,9	5
14,6	1	16,6	3	27,6	4	31,1	5	36,6	5

1) Για scatterplot: Graphs/Scatterplot/simple, Y axis: y, X axis: f. Για Boxplot: Graphs/Boxplot/Simple: Variable:y, Category Axis: f





Παρατηρούμε ότι ο 1ος τύπος αυτοκινήτου φαίνεται να έχει να έχει την μικρότερη κατανάλωση βενζίνης, ενώ ο 5ος τύπος φαίνεται να έχει την μεγαλύτερη κατανάλωση. Για να διαπιστώσουμε όμως αν αυτή η διαφορά είναι σημαντική θα πρέπει να προχωρήσουμε στην ανάλυση διακύμανσης.

2) Για να ελέγξουμε αν οι διασπορές των παρατηρήσεων είναι ίδιες στις περιπτώσεις 1, 2, 3 με α = 0.1 θα χρησιμοποιήσουμε και πάλι το τεστ Levene που περιγράφεται στην λύση της προηγούμενης άσκησης. Επιλέγουμε Analyze/compare means/one-way ANOVA, Dependent list: y Factor: f, : Options: check Homogeneity of variance από όπου παίρνουμε

Test of Homogeneity of Variances

Y

Levene Statistic	df1	df2	Sig.
,684	4	45	,607

Εφ'όσον το p-value = 0.607 δεχόμαστε ότι οι διασπορές των παρατηρήσεων είναι ίδιες στις 5 περιπτώσεις.

3) Εφαρμόζουμε το μοντέλο

$$Y_{ij} = \mu_i + \varepsilon_{ij},\ i = 1,2,\ldots,5, \quad j = 1,2,\ldots,n_i,$$

όπου ε_{ij} ανεξάρτητες τ.μ. που ακολουθούν N(0,σ^2). Στο SPSS: Analyze/compare means/one-way ANOVA, Dependent list: y Factor: f, Options: Check Descriptive λαμβάνεται ο πίνακας:

Descriptives

Y

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	11	17,673	2,940	,886	15,698	19,648	13,0	21,7
2	8	27,513	2,436	,861	25,476	29,549	23,5	30,1
3	10	22,080	3,610	1,142	19,497	24,663	16,6	27,1
4	8	25,975	4,039	1,428	22,598	29,352	20,3	32,8
5	13	32,292	3,232	,896	30,339	34,246	27,9	38,5
Total	50	25,258	6,238	,882	23,485	27,031	13,0	38,5

ο οποίος περιέχει τις εκτιμήσεις και τα δ.ε. 95% για τα $\mu_1,\ldots,\mu_5$ χρησιμοποιώντας internal variance estimates. Επίσης λαμβάνεται ο πίνακας

ANOVA

Y

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1421,931	4	355,483	32,993	,000
Within Groups	484,851	45	10,774		
Total	1906,782	49			

από όπου λαμβάνεται ότι $s^2 = MSE = 10.774$ και άρα τα δ.ε. 1−α για τα μ_i με pooled εκτίμηση της διασποράς είναι

$$\overline{Y}_{1\bullet} \pm \frac{s}{\sqrt{n_1}} t_{n_T - r}(\frac{a}{2}) = 17.673 \pm \frac{\sqrt{10.774}}{\sqrt{11}} t_{45}(0.025) = (15.6798, 19.6662)$$

$$\overline{Y}_{2\bullet} \pm \frac{s}{\sqrt{n_2}} t_{n_T - r}(\frac{a}{2}) = 27.513 \pm \frac{\sqrt{10.774}}{\sqrt{8}} t_{45}(0.025) = (25.1758, 29.8502)$$

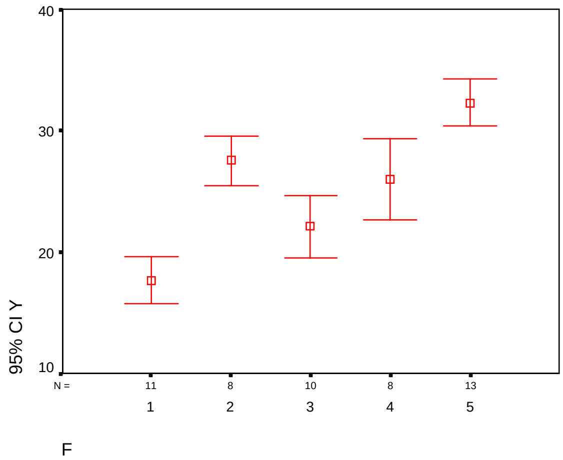
$$\overline{Y}_{3\bullet} \pm \frac{s}{\sqrt{n_3}} t_{n_T - r}(\frac{a}{2}) = 22.080 \pm \frac{\sqrt{10.774}}{\sqrt{10}} t_{45}(0.025) = (19.9895, 24.1705)$$

$$\overline{Y}_{4\bullet} \pm \frac{s}{\sqrt{n_4}} t_{n_T - r}(\frac{a}{2}) = 25.975 \pm \frac{\sqrt{10.774}}{\sqrt{8}} t_{45}(0.025) = (23.6378, 28.3122)$$

$$\overline{Y}_{5\bullet} \pm \frac{s}{\sqrt{n_5}} t_{n_T - r}(\frac{a}{2}) = 32.292 \pm \frac{\sqrt{10.774}}{\sqrt{13}} t_{45}(0.025) = (30.4585, 34.1255)$$

όπου $t_{45}(0.025) = IDF.T(0.975, 27) \approx 2.014$. Τα παραπάνω δ.ε. μπορούν να εξαχθούν αυτόματα και από τη διαδικασία Linear Regression (βλ. (7)) παρακάτω.

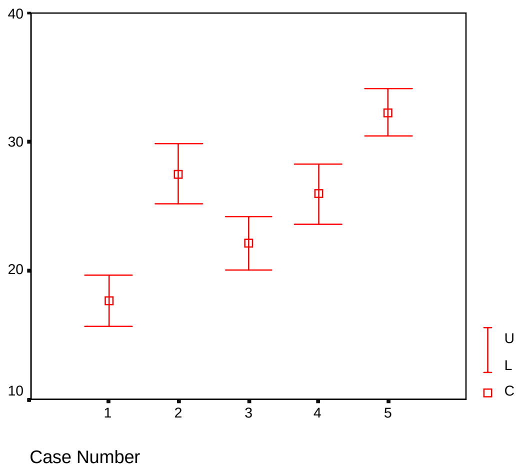
Για τα γραφήματα των δ.ε. 95% για τα μ_i (με τα individual variance estimates) επιλέγουμε Graphs/error bars/simple, variable: y, Category Axis: f (c.i. for mean, level: 95%) από όπου λαμβάνουμε το γράφημα



Για το αντίστοιχο γράφημα με τα pooled estimates φτιάχνουμε τρεις μεταβλητές που περιέχουν τα συγκεκριμένα δ.ε. (με τα κεντρικά σημεία) ως εξής:

L	*C*	*U*
15.6798	17.673	19.6662
25.1758	27.513	29.8502
19.9895	22.080	24.1705
23.6378	25.975	28.3122
30.4585	32.292	34.1255

και επιλέγουμε Graphs/High-Low/Simple High-Low, check values of individual cases, High:U, Low: L, close: C από όπου λαμβάνουμε το γράφημα με τα δ.ε. 95% για τα μ_i (pooled variance estimates):



4) Ο έλεγχος της υπόθεσης H_0: $\mu_1 = \ldots = \mu_5$ έναντι της H_1: όχι $\mu_1 = \ldots = \mu_5$ γίνεται χρησιμοποιώντας το *F - test* στον πίνακα ANOVA. Παρατηρούμε ότι το αντίστοιχο p-value είναι 0.00 ($F^* = 32.993$) οπότε απορρίπτουμε ότι $\mu_1 = \ldots = \mu_5$ σε ε.σ. 1% (η μέση τιμή δεν είναι ίση και στις πέντε ομάδες).

5) Επιλέγουμε Analyze/compare means/one-way ANOVA, Dependent list: y Factor: f, : Post Hoc: LSD απ’ όπου λαμβάνουμε:

Multiple Comparisons

Dependent Variable: Y
LSD

(I) F	(J) F	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
1	2	-9,840*	1,525	,000	-12,912	-6,768
	3	-4,407*	1,434	,004	-7,296	-1,519
	4	-8,302*	1,525	,000	-11,374	-5,230
	5	-14,620*	1,345	,000	-17,328	-11,911
2	1	9,840*	1,525	,000	6,768	12,912
	3	5,433*	1,557	,001	2,297	8,568
	4	1,537	1,641	,354	-1,768	4,843
	5	-4,780*	1,475	,002	-7,751	-1,809
3	1	4,407*	1,434	,004	1,519	7,296
	2	-5,433*	1,557	,001	-8,568	-2,297
	4	-3,895*	1,557	,016	-7,031	-,759
	5	-10,212*	1,381	,000	-12,993	-7,431
4	1	8,302*	1,525	,000	5,230	11,374
	2	-1,537	1,641	,354	-4,843	1,768
	3	3,895*	1,557	,016	,759	7,031
	5	-6,317*	1,475	,000	-9,288	-3,347
5	1	14,620*	1,345	,000	11,911	17,328
	2	4,780*	1,475	,002	1,809	7,751
	3	10,212*	1,381	,000	7,431	12,993
	4	6,317*	1,475	,000	3,347	9,288

*. The mean difference is significant at the .05 level.

οπότε το 95% δ.ε. για τη διαφορά $\mu_5 - \mu_1$ είναι (11.911, 17.328) και το 95% δ.ε. για τη διαφορά $\mu_4 - \mu_2$ είναι (–4.843, 1.768). Για τον έλεγχο H_0: $\mu_4 = \mu_2$ παρατηρούμε ότι το αντίστοιχο p-value είναι 0.354 > 0.01 και άρα μπορούμε να πούμε ότι το μ_4 δεν διαφέρει σημαντικά από το μ_2 (ε.σ. 1%).

6) Επιλέγουμε Analyze/compare means/one-way ANOVA, Dependent list: y Factor: f, : Post Hoc: Tukey, Sheffe, Bonferroni απ' όπου λαμβάνουμε:

	(I) F	(J) F	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Tukey HSD	1	2	-9,840	1,525	**,000**	-14,174	-5,506
		3	-4,407	1,434	**,028**	-8,483	-,332
		4	-8,302	1,525	**,000**	-12,636	-3,968
		5	-14,620	1,345	**,000**	-18,441	-10,799
	2	1	9,840	1,525	**,000**	5,506	14,174
		3	5,433	1,557	**,009**	1,008	9,857
		4	1,537	1,641	,881	-3,126	6,201
		5	-4,780	1,475	**,018**	-8,971	-,589
	3	1	4,407	1,434	**,028**	,332	8,483
		2	-5,433	1,557	**,009**	-9,857	-1,008
		4	-3,895	1,557	,108	-8,319	,529
		5	-10,212	1,381	**,000**	-14,135	-6,289
	4	1	8,302	1,525	**,000**	3,968	12,636
		2	-1,537	1,641	,881	-6,201	3,126
		3	3,895	1,557	,108	-,529	8,319
		5	-6,317	1,475	**,001**	-10,508	-2,126
	5	1	14,620	1,345	**,000**	10,799	18,441
		2	4,780	1,475	**,018**	,589	8,971
		3	10,212	1,381	**,000**	6,289	14,135
		4	6,317	1,475	**,001**	2,126	10,508

Scheffe	1	2	-9,840	1,525	**,000**	-14,738	-4,941
		3	-4,407	1,434	,067	-9,013	,199
		4	-8,302	1,525	**,000**	-13,201	-3,404
		5	-14,620	1,345	**,000**	-18,938	-10,301
	2	1	9,840	1,525	**,000**	4,941	14,738
		3	5,433	1,557	**,026**	,432	10,433
		4	1,537	1,641	,926	-3,734	6,809
		5	-4,780	1,475	**,047**	-9,517	-4,258E-02
	3	1	4,407	1,434	,067	-,199	9,013
		2	-5,433	1,557	**,026**	-10,433	-,432
		4	-3,895	1,557	,200	-8,896	1,106
		5	-10,212	1,381	**,000**	-14,647	-5,778
	4	1	8,302	1,525	**,000**	3,404	13,201
		2	-1,537	1,641	,926	-6,809	3,734
		3	3,895	1,557	,200	-1,106	8,896
		5	-6,317	1,475	**,003**	-11,055	-1,580
	5	1	14,620	1,345	**,000**	10,301	18,938
		2	4,780	1,475	**,047**	4,258E-02	9,517
		3	10,212	1,381	**,000**	5,778	14,647
		4	6,317	1,475	**,003**	1,580	11,055
Bonferroni	1	2	-9,840	1,525	**,000**	-14,342	-5,337
		3	-4,407	1,434	**,036**	-8,641	-,173
		4	-8,302	1,525	**,000**	-12,805	-3,800
		5	-14,620	1,345	**,000**	-18,589	-10,650
	2	1	9,840	1,525	**,000**	5,337	14,342
		3	5,433	1,557	**,011**	,836	10,029
		4	1,537	1,641	1,000	-3,308	6,383
		5	-4,780	1,475	**,022**	-9,134	-,426
	3	1	4,407	1,434	**,036**	,173	8,641
		2	-5,433	1,557	**,011**	-10,029	-,836
		4	-3,895	1,557	,161	-8,491	,701
		5	-10,212	1,381	**,000**	-14,288	-6,136
	4	1	8,302	1,525	**,000**	3,800	12,805
		2	-1,537	1,641	1,000	-6,383	3,308
		3	3,895	1,557	,161	-,701	8,491
		5	-6,317	1,475	**,001**	-10,672	-1,963
	5	1	14,620	1,345	**,000**	10,650	18,589
		2	4,780	1,475	**,022**	,426	9,134
		3	10,212	1,381	**,000**	6,136	14,288
		4	6,317	1,475	**,001**	1,963	10,672

Η ομαδοποίηση των μέσων είναι:

Y

	F	N	Subset for alpha = .05			
			1	2	3	4
Tukey HSD[a,b]	1	11	17,673			
	3	10		22,080		
	4	8		25,975	25,975	
	2	8			27,513	
	5	13				32,292
	Sig.		1,000	,086	,841	1,000
Scheffe[a,b]	1	11	17,673			
	3	10	22,080	22,080		
	4	8		25,975	25,975	
	2	8			27,513	27,513
	5	13				32,292
	Sig.		,087	,167	,899	,051

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 9,656.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

7) το μοντέλο μας μπορεί εναλλακτικά να γραφεί στη μορφή $Y = \mathbf{X}\beta + \varepsilon$ όπου

$$Y=\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \vdots \\ \vdots \\ Y_{r1} \\ \vdots \\ Y_{rn_r} \end{bmatrix}, \quad \mathbf{X}=\begin{bmatrix} 1 & & \\ \vdots & & \\ 1 & & \\ & \ddots & \\ & & \ddots \\ & & 1 \\ & & \vdots \\ & & 1 \end{bmatrix}, \beta=\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_r \end{bmatrix} \varepsilon=\begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \vdots \\ \vdots \\ \varepsilon_{r1} \\ \vdots \\ \varepsilon_{rn_r} \end{bmatrix}$$

συνεπώς εκτελώντας Analyze/Regression/Linear/Dependent:y, Independent:X_1,X_2,X_3,X_4,X_5 options: do not include constant in equation, Statistics: check confidence intervals λαμβάνουμε τον πίνακα:

Coefficients[a,b]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	X1	17,673	,990	,319	17,857	,000	15,679	19,666
	X2	27,512	1,161	,423	23,707	,000	25,175	29,850
	X3	22,080	1,038	,380	21,272	,000	19,989	24,171
	X4	25,975	1,161	,400	22,382	,000	23,638	28,312
	X5	32,292	,910	,633	35,471	,000	30,459	34,126

a. Dependent Variable: Y

b. Linear Regression through the Origin

Τα παραπάνω δ.ε. είναι ακριβώς αυτά που υπολογίσαμε παραπάνω χρησιμοποιώντας pooled variance estimates.

Τέλος, ο πίνακας ANOVA προκύπτει (σύμφωνα με σχόλια στην προηγ. Άσκηση) από τη διαδικασία Analyze/ Regression/ Linear/ Dependent: y, Independent:X_2, X_3, X_4, X_5 options: include constant in equation:

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1421,931	4	355,483	32,993	,000[a]
	Residual	484,851	45	10,774		
	Total	1906,782	49			

a. Predictors: (Constant), X5, X4, X2, X3

b. Dependent Variable: Y